2005

Contribution towards a fast stereo dense matching.

Rabihaa Zein
University of Windsor

CONTRIBUTION TOWARDS A FAST STEREO DENSE MATCHING

by

Rabihaa Zein

A Thesis
Submitted to the Faculty of Graduate Studies and Research
through Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

2005

Your file Votre référence
ISBN: 0-494-09787-6
Our file Notre référence
ISBN: 0-494-09787-6

Abstract

Stereo matching is important in the area of computer vision as it is the basis of the reconstruction process. Many applications require 3D reconstruction such as view synthesis, robotics... The main task of matching uncalibrated images is to determine the corresponding pixels and other features where the motion between these images and the camera parameters is unknown.

Although some methods have been carried out over the past two decades on the matching problem, most of these methods are not practical and difficult to implement. Our approach considers a reliable image edge features in order to develop a fast and practical method. Therefore, we propose a fast stereo matching algorithm combining two different approaches for matching as the image is segmented into two sets of regions: edge regions and non-edge regions. We have used an algebraic method that preserves disparity continuity at the object continuous surfaces. Our results demonstrate that we gain a speed dense matching while the implementation is kept simple and straightforward.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The purpose of computer vision is to program a computer to 'understand' a scene or features in an image. Computer Vision refers to the application of human vision techniques to a computer, teaching the computer to see. As humans, we rely on two perspectives of a scene to perceive 3D data. Similarly, many computer vision applications (such as multimedia, robotics, view synthesis...) need at least two images of a scene to recover 3D structure. Recovering 3D information from a pair of stereo images is essentially the correspondence problem. However, one of the most direct way of achieving the 3D data from image data is stereo vision.

1.1 Stereoscopic vision

In computer vision, stereoscopic (stereo for solid) vision is used when no information is known about a scene and the analysis of a single image of that scene is not enough to recover 3D information. It is commonly known that stereopsis is the primary way for humans to perceive depth. Although we can still interact very well with our environment with one eye and do very highly skillful tasks by using other visual cues such as occlusion and motion, the resultant effect of the absence of stereopsis is that the relative depth information between objects is essentially lost. Similarly, at least two

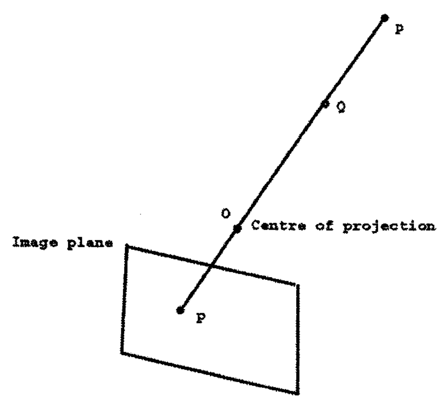


Figure 1.1: Projection of 3D point on the image plan.

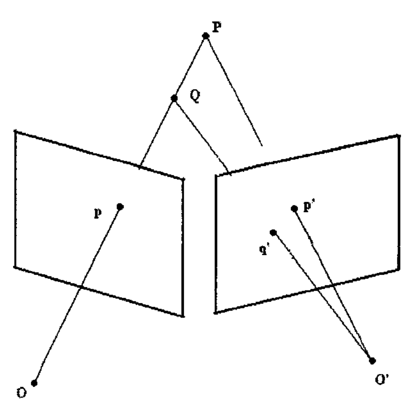


Figure 1.2: At least two images are required in stereo vision.

images are needed to recover the 3D structure using a stereo vision system. Therefore, stereo vision is concerned with the analysis and study of the imaging process using two cameras observing the same scene. An image is obtained by the projection of a 3D structure P onto a 3D plane p (Figure 1.1). Suppose that we try to find the depth of the point P in space that projects on the image point P' in the image. Information like absolute scale and depth is lost when the scene is projected onto an image plane. In fact, there are an infinite number of points along the ray formed by OP could have projected at p, which makes direct computation of the depth of P from a single image impossible. However, with the help of another image (Figure 1.2), the depth of P can be easily calculated given its projection p in the left image and p_1' in the right image.

1.2 Matching Problem of Stereo Images

Classical methods solve the problem of obtaining the depth (3D reconstruction) of a particular pixel image in three steps: camera calibration, matching the features in different images, and then 3D reconstruction.

1. Calibrate both cameras so that the geometrical relationships between the 3D space and the two cameras become known (calibration process).

2. Identify the image point that represents the same scene point in the other image (matching process).

3. Calculate the depth of the selected location based on the location difference of the corresponding points and camera positions (reconstruction process).

The classical solution is not realistic because of the burden of the calibration process. A slight change in the camera position, orientation, or focus will require the re-calibration of that camera. In the early-nineties, self-calibration (camera auto-calibration) was introduced and based on inferring information from the analysis of a sequences of images, provided a more practical alternative to the classical approach. However, without the calibration process, the classical solution represent two problems: correspondence and reconstruction.

Given two images formed in the retinal planes, we want to solve two problems:

- For a point m in the first image, determine which point m' in the second image that corresponds to. The term correspond means that they are the images of the same physical point M in the scene. This is what is commonly known as the correspondence problem. The problem of correspondence is also known as the matching problem. Once matching is achieved, calculation of depth becomes a straightforward geometrical problem

- Given two corresponding points m and m', compute the 3-D coordinates of M relative to some global reference frame. This is known as the reconstruction problem.

However, the problem of point matching is to find the correspondence of pixels between two images.

1.3 Matching of Uncalibrated Images

The class of a matching technique is categorized according to the scope of its solution. A large amount of work has been done to solve the dense matching problem of uncalibrated image. Because stereo vision is inherently complicated and noise sensitive, classical approaches either were limited to dense matching and sparse matching

- Dense matching approach aims at matching each image pixel which is in fact the center of a small window of points in the first image that is compared with same sized windows in the second image. It doesn't require information about camera parameters and therefore it is a natural choice for dense matching of uncalibrated images. It achieves a dense matching result for the stereo image pair.

- Sparse matching approach restricts the search for correspondences to a sparse set of features such as edges, lines and dots. Sparse matching approach provides sparser depth data which is locally more accurate and globally more reliable.

When images are uncalibrated, the motion between them and the camera parameters are unknown and the only information available is the raw images themselves. The basic relationship between a pair of images is the correspondence between their pixels which considers a challenge problem of dense matching. Therefore, the objective of this research is as follows:

- Review of the existing methods for dense matching of uncalibrated images
- Design and implement a hybrid matching algorithm that integrate edge feature of uncalibrated stereo images in order to obtain a fast dense matching.
- In order to match the non-edge areas of the image, we implement an algebraic method that preserves disparity continuity at the object continuous surfaces.
- Evaluate our proposed method using experiments for indoor and outdoor scenes.

1.4 Organization of the thesis

This thesis is presented in the following order: Chapter 2 presents the required geometrical background to understand the problem related to stereo matching in computer vision. Chapter 3 describes the previous work in the area of stereo vision and the work is divided into two category, sparse matching techniques and dense matching techniques for obtaining matching of uncalibrated images. Chapter 4 proposes a fast dense matching algorithm which integrates edge features of the image. The matching is carried out separately for edge regions and non-edge regions. We used an algebraic method that restricts the search area in order to match the non-edge regions of the image. Our experimental results demonstrate the capabilities and efficiency of our hybrid method. Chapter 5 is a conclusion and suggestions for future work.

Chapter 2

Background of Matching

Before outlining the principles of dense matching, it is important to know how images are formed. First, the image formation is given, the camera model and the camera projection matrix is introduced. We present some constraints that are sometimes used to reduce the matching process. Epipolar constraints was identified as one major constraint used by most matching method as it has the advantage of being available without calibrating the cameras. Dense matching represents a fundamental problem in computer vision. First solution for image matching have been suggested already in the late fifties [10]. Since then a steady increase in the interest for image matching has occurred.

2.1 Image Formation: a geometric point of view

The image formation process is the process by which a 3D-representation of a scene is reduced to a 2D-representation of that scene. The projection of light rays onto the retina presents our visual system with an image of the world that is inherently two-dimensional, yet we are able to interact with the three-dimensional world, even in situations new to us, or with objects unknown to us. An image is obtained with a CCD camera, stands for charge coupled device. However, the image is the result of a geometric transformation, which takes a 3D description of a scene and changes it into a 2D description.

2.2 Camera Model

Camera model is a mathematical approximation of the image formation process in real world cameras. When a light ray is reflected off an object surface and passes through the pinhole lens of a camera onto the sensor array, an image is obtained. This image is a projection of the object surface; it is also called the image plane. A point on the image plane is a pure perspective projection of a point in the scene, while an object in the scene is represented by a collection of pixels in the image.

2.2.1 Pinhole Model

A camera is usually described using the *pinhole model* which is also known as the full perspective model. As shown in (Figure 2.1), a camera is attached to a 3D reference frame with origin O, called center of projection, and three axes Ox, Oy and Oz. The line through O and perpendicular to the image plane is the optical axis OZ; Let $P = [X_w, Y_w, Z_w]^T$ and $p = [x, y, z]^T$, this reference frame is called the camera reference, the focal length f of the camera which is the distance between the center of projection and the retinal plane. The camera reference frame is located within another frame called the world reference frame or a scene reference frame. The image coordinate system is also called the pixel coordinate system. The parameters linking the world coordinate system to image pixels are:

- Intrinsic Parameters: there are five intrinsic camera parameters: the parameters that are needed to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame.

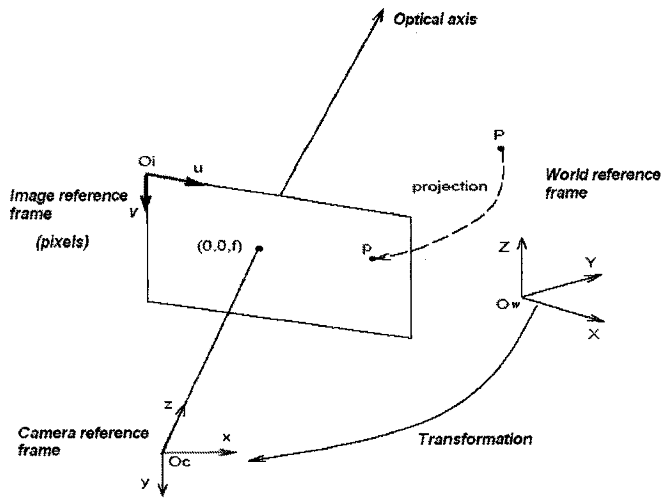


Figure 2.1: Image Formation.

- $f_x = f/s_x$, is the length in effective horizontal pixel size units (s_x, s_y).
- $a = s_y/s_x$, aspect ratio
- (o_x, o_y), are the image center coordinates
- k_1, radial distortion coefficient

- Extrinsic Parameters: there are six intrinsic camera parameters: three are the position of the center of projection, and three are for the orientation of the image plane coordinate frame. The parameters that define the location and orientation of the camera reference frame with respect to a known world reference frame
  - R, the 3 x 3 rotation matrix
  - T, the 3d translation vector.

However, estimating the intrinsic and extrinsic parameters is carried out through the recovery of the parameters of the projection matrix M. Calibration allows the recovery of the parameters that governs the image formation process 3D to 2D, with the ultimate goal of obtaining the 3D model based on 2D information (pixel coordinates).

Geometrical Transformation of a 3D Point into a Pixel

1. 3D scene : the scene coordinates are usually not written with respect to the camera reference frame. The transformation has to take into account the change of coordinate system ($World \rightarrow Camera$). A world point $P = (X_w, Y_w, Z_w)$ undergoes a 3D transformation (a rotation R and a translation t) to be fit into the camera reference frame :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

A compact version of the above will be: $P' = DP$ Where D is a 4×4 matrix representing the extrinsic parameters

$$D = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_y \end{pmatrix}$$

P and p are the homogeneous coordinates in the scene and camera reference frames respectively.

2. The 3D-2D Transformation: a perspective projection converts a 3D point $P(X, Y, Z)$ to a 2D point $p = (x, y, z)$.
In this case we can write:

$$\begin{cases} \mathrm{x} = f/\mathrm{Z} \\ \mathrm{y} = f/\mathrm{Z} \\ \mathrm{z} = f \end{cases}$$

In fact, this relation is in fact a projective transformation :

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

A compact version of the above expression will be :

$$p = \lambda I P$$

where I is a 3×4 matrix representing a perspective projection and, λ is a scale factor for equality in the projective space.

3. The 2D-2D Transformation: 2D point is transformed from the camera reference frame to the image reference frame. Coordinates in this reference frame correspond to the pixel positions (u and v) of the point. Other parameters are involved in this transformation: u_0, v_0 and w_0 are the coordinates of the center of projection of two scaling factors k_u and k_v (pixels/mm) represent respectively the horizontal and vertical pixel dimensions, since pixels in real world cameras are rarely square. The transformation now writes :

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.1}$$

where A represents the matrix of intrinsic parameters, namely, α_u, α_v, α_{u_0}, and α_{v_0}. This is an Affine transformation made of a scale change (u and v), and a translation. Matrix A is a linear transformation from the projective space to the projective plane.

α_u and α_v are the 2 scale factors along the image Ox and Oy directions respectively. The perspective projection that transforms space points defined on the world coordinate system, into image points defined in the image coordinate system, can be described by a 3×4 matrix, often denoted by M, where $M = AID$:

D represents the matrix from world to camera reference frame (extrinsic parameters)

I represents $3D - 2D$ projection

A represents $2D - 2D$ transformation (intrinsic parameters).

2.3 Disparity and Parallel Stereo Images

- The parallel camera case is a special stereo camera configuration in which the two retinal planes are horizontally displaced and are coplanar in space, and the two cameras have identical focal length. As shown in Figure 2.2, given a scene point M and its two projection points m of coordinates (u, v) and m' of coordinates (u', v'), the disparity value d is defined as $d = u' - u$. Note that $v = v'$ as there is no vertical parallax between the two cameras. Therefore, The value of d shows how much the position of the pixel has moved between m and m ', this is known as the disparity of the pixel. The results are shown in a disparity map. The depth measure z of M is related to the disparity value d.

- In the general case, given two corresponding points that have coordinates (u, v) and (u', v') projected by a scene point M in a general stereo camera configuration, the disparity measure between m and m' is in fact a 2-vector instead of a real number. In this case, the depth value z of M is the perpendicular distance of M from the baseline t and disparity d is a real number that measures the inverse of

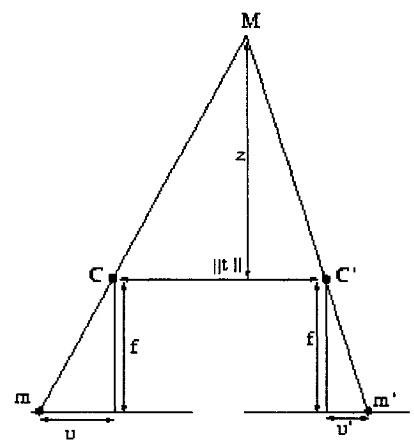


Figure 2.2: Disparity for parallel cameras.

z.

- In parallel camera configurations, the epipolar lines coincide with the horizontal scan lines (shown in Figure 2.3), and the epipoles are at infinity, c and c' are the projection centers of two cameras. As will be seen in the following chapters, stereo matching is greatly simplified for parallel cameras.

- In general stereo configurations (shown in Figure 2.4), the epipolar lines in at least one retinal plane intersect at a point called the epipole. Given a pair of stereo images, rectification determines a transformation of each image. This step consists of transforming the images so that the epipolar lines are aligned horizontally. The importance of rectification is to reduce the correspondence problem from 2-D search to just 1-D search Loop (Zhang [17]).

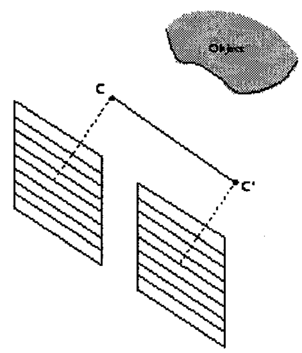


Figure 2.3: Parallel stereo image configuration.

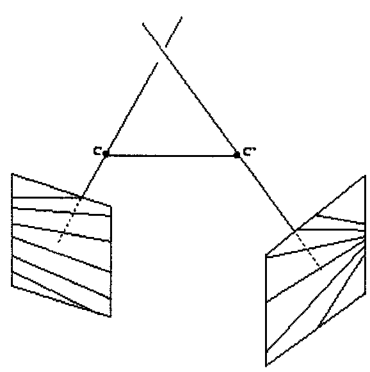


Figure 2.4: General stereo configurations.

2.4 Stereo Matching Correlations

The most common way of stereo matching is to use a correlation function: measure of match. There are many different correlation scores based on different statistics of the neighborhood. The most common correlation measures is ZNCC correlation which stands for Zero-mean Normalized Cross-Correlation and we will use later in our work because it is very robust against many types of image distortion and noise. The basic formula for cross correlation can be derived from the sum of squared differences of two pixel's neighborhoods. The calculation of correlation measures is computationally expensive and do not provide a unique global maximal. Some of the more common measures are listed in Table 2.1, these correlation measures vary in complexity. The calculation of correlation measures is computationally expensive. Measures such as SSD and SAD are simple to compute, but may not be robust. Depending on the specific measure, the goal may be either minimize or maximize the correlation score.

ZNCC has evolved from the cross correlation (CC) function. From ZNCC equation one can note that the denominator is the multiplication and square root of two parts: these parts compute the square difference of a pixel and its associated template's average value $\overline{I}$ and $\overline{I'}$. Here $\overline{I}$ denotes the average intensity of the target window of a candidate match, and $\overline{I'}$ the average intensity of the reference window of the pixel to be matched. The averages are calculated only once, and can be retrieved each time needed. Therefore, the total computation cost is considerably reduced.

The following sections introduce several of approaches for the dense matching of uncalibrated images. Each solution differs in the specific search strategy implemented. However, in order to evaluate candidate matches, each method relies on a correlation measure.

Name	Formula
Sum of Squared Differences	$SSD((x,y),(x\prime,y\prime)) = \sum_{i=-n}^{m}\sum_{j=-m}^{m}(I(x+i,y+j)-I\prime(x\prime+i,y\prime+j))^2$
Sum of absolute Differences	$SAD((x,y),(x\prime,y\prime)) = \sum_{i=-n}^{m}\sum_{j=-m}^{m}\lvert I(x+i,y+j)-I\prime(x\prime+i,y\prime+j)\rvert$
Cross Correlation	$CC((x,y),(x\prime,y\prime)) = \sum_{i=-n}^{m}\sum_{j=-m}^{m}I(x+i,y+j).I\prime(x\prime+i,y\prime+j)$
Zero Mean Normalized Cross Correlation	$ZNCC((x,y),(x\prime,y\prime)) = \frac{\sum_{i=-n}^{m}\sum_{j=-m}^{m}(I(x+i,y+j)-\overline{I}).(I\prime(x\prime+i,y\prime+j)-\overline{I\prime})}{\sqrt{\sum_{i=-n}^{m}\sum_{j=-m}^{m}(I(x+i,y+j)-\overline{I})^2.\sum_{i=-n}^{m}\sum_{j=-m}^{m}(I(x\prime+i,y\prime+j)-\overline{I\prime})^2}}$

Table 2.1: Common Correlation Functions.

2.5 Stereo Matching Constraints

Stereo matching process is a very difficult search procedure. In order to minimize false matches, some matching constraints must be imposed.

The epipolar geometry can be used to reduce this problem from a 2D search to a 1D search. Epipolar geometry was identified as one major constraint used by most matching methods.

- Epipolar Constraint: in stereo vision, two views geometry is subject to follow the epipolar constraint which constraint the correspondence $p\prime$ of a point p to lie on the epipolar line. An epipolar line is defined by the intersection of an epipolar plane and an image. Therefore, the correct match of p must lie on this corresponding epipolar line in the right image. When images are uncalibrated, the motion between them and the camera parameters are not known. Epipolar geometry has the advantage of being available without calibrating the cameras. In computer vision, the most common way to describe the epipolar geometry is by means of a 3×3 *matrix* called the fundamental matrix.

  Figure 2.5 shows that epipolar geometry exists between images of a two camera system. The above relationship illustrates that the projection of a point P in space into two different 2D points p on the first view and $p\prime$ on the second view. Suppose there are two pinhole cameras, with their projection centers O and $O\prime$ respectively, and two image planes. The projection of O on the first image and $O\prime$ on the second image are denoted respectively by e and e', where e and e' are called epipoles. Points P, O and O' form a plane are called the epipolar plane. An epipolar line is defined by the intersection of an epipolar plane and an image.

The line defined in the first image by e and p is an epipolar line, and its corresponding epipolar line in the second image is defined by e' and p' (denoted by lep and $l'ep$ respectively). In stereo vision, two views geometry is subject to the epipolar constraint [16] which constraint the corresponding point p' of a point p to lie on the epipolar line $l'ep$. The mapping between points in one image and epipolar lines in the other can be established by estimating two important matrices: the essential matrix E and the fundamental matrix F.

- Essential Matrix E: establishes a natural link between the epipolar constraint and the extrinsic parameters of the stereo system. E is the mapping between points and epipolar lines we are looking for. Satisfies the equation:

$$p\prime^T E p = 0 \tag{2.2}$$

  where p is in camera coordinates

- Fundamental Matrix F: in the uncalibrated case, the interpretation of the epipolar constraints yields another constraint which is the Fundamental matrix F[18] that contains the geometric information which relates a couple of stereo images. The fundamental matrix F is a generalization of the essential matrix E and the relationship between these two matrices was given by Luong [18]. F contains the geometric information which relates a couple of stereo images (eq. 1.1):

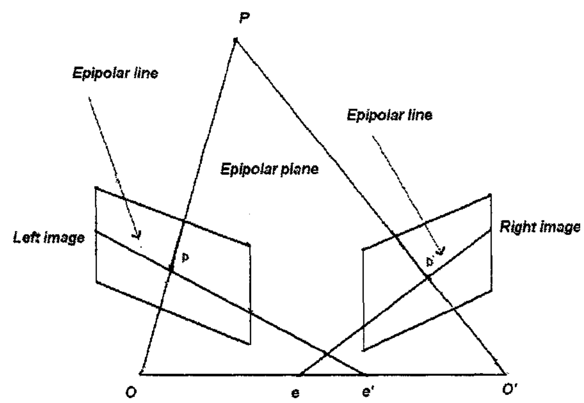


Figure 2.5: Epipolar geometry.

$$p\prime^T Fp = 0 \tag{2.3}$$

The difference from the Essential Matrix is that F is defined in terms of pixel coordinates, while E is defined in terms of camera coordinates. Relationship between E and F is $F = (M^{-1})^T EM^{-1}$ where M are the matrices of the left and right intrinsic parameters. $Fp = (a, b, c)$ represents the coefficients of the corresponding epipolar line in the right image, on which $p\prime$ should be located. When fundamental matrix F is known, the matching process can be simplified from a 2D to a 1D problem. F can be computed when given 8 or more matched oints in a pair of uncalibratd images. This simple algorithm is known as the eight -point algorithm for calculating F. Some new methods [11] for calculating the fundamental matrix F are very reliable, there is no guarantee that F is accurate each time.

- Continuity Constraint: the cohesiveness of matters suggests that the disparity of the matches should vary smoothly almost everywhere over the image. This constraint fails at discontinuities of depth, for depth discontinuities cause an

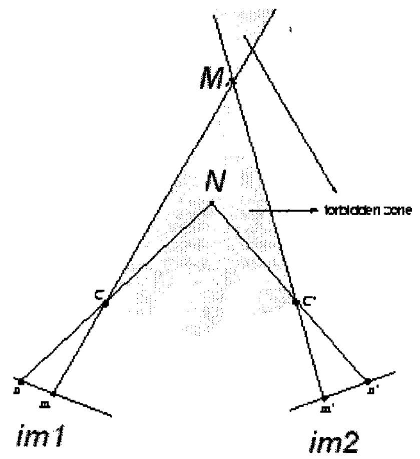


Figure 2.6: The ordering constraint.

abrupt change in disparity.

- Uniqueness Constraint: a given pixel or feature from one image can match no more than one pixel or feature from the other image. This constraint can also fail if transparent objects are present in the scene. Furthermore, given a pixel or feature m in one image, its 'corresponding' pixel or feature may be occluded in the other image. In this case, no match should be assigned to m.

- Order Constraint: the order constraint states that the order of points along the according epipolar line is preserved. If m is to the left of n then m' should also be to the left of n' and vice versa. That is, the ordering of features is preserved across images. The ordering constraint fails if a given 3D point N falls onto the forbidden zone of another 3D point M. In im1, m is on the right of n, but in im2, this ordering is reserved (Figure 2.6).

Unlike all the other constraints, the epipolar constraint would never fail and could be applied reliably once the epipolar geometry is known.

2.6 Stereo Matching Techniques

Traditional matching algorithm involves into a search to select a set of candidate matches and a measure to evaluate them. Basically, two types of matching techniques are considered: Dense matching technique considers only the intensity of the pixels and Sparse matching method establishs correspondences between extracted features of images.

- Sparse-based Techniques: in the sparse-based techniques, features are first extracted from the images and the matching process is applied to the features. The detection of image features can be viewed as a processing stage for matching. Features are extracted in each image individually prior to matching them. Instead of using correlation as similarity measurement, corresponding elements are given by the most familiar feature pair using other criteria.

  Most methods narrow the number of possible features with which to match by constraints: geometric constraints, analytical constraints. The image pair is first preprocessed by an operator so as to extract the features that are stable under the change of viewpoint, the matching process is then applied to the attributes associated with the detected features. The obvious question here is what type of features that one should use? Edge elements and corners have been widely used in many stereo vision work and are easy to detect, but may suffer from occlusion; line and curve segments require extra computation time, but are more robust against occlusion (they are longer and so are less likely to be completely occluded). Higher level image features such as circles, ellipses, and polygonal regions have also been used as features for stereo matching, these features are, however, restricted to images of indoor scenes.

  Sparse-based techniques have the advantage of being sufficient and robust to noise but they are difficult to implement.

- Dense Matching Techniques: it is a correlation-based techniques. Each image

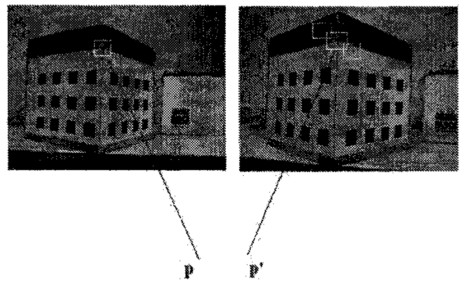


Figure 2.7: Correlation.

point to be matched is in fact the center of a small window of points in the first image that is compared with same sized windows in the second image. Traditional matching algorithm involves into a search to select a set of candidate matches and a measure to evaluate them. The goal of the search strategy is to employ heuristics to find a minimal set containing good candidate matches. This may include the application of constraints to the image or the use of a numerical approach to find the optimal candidate. Once the candidate set has been found, a correlation function (see section 2.5) is used to evaluate the candidates. An example is shown in Figure 2.7, to match a point p in the left image, a small window is then compared with same sized windows in the right image for each pixel in the search area. Each comparison produces a correlation score using certain correlation functions. The candidate match $p\prime$ shall be associated with the window that maximizes the similarity function. However, the correspondences between images are determined based on the similarities of the pixels gray value using the image template windows.

Correlations techniques are sensitive to lighting changes and computationally expensive but they are easy to implement.

2.7 Conclusion

The geometry of stereo vision system and some important concepts was introduced in this chapter. We presented the approaches for dense matching, stereo matching correlations and constraints. ZNCC considered the most common correlation measures, as it is more robust about noise in the image. The importance of existing constraints was introduced and in particular the epipolar constraint which reduces the search of the matching process from a two-dimensional search to a one-dimensional search. The next chapter presents a review of previous techniques that were carried out over the last two decades on the matching problem.

Chapter 3

Existing Stereo Matching Techniques

Approaches to the correspondence problem can be broadly classified into two categories: the dense matching or area-matching and the sparse matching techniques. Sparse matching method is concerned only with identifying and matching interest points. These points include corners, junctions and dots. This problem is relatively simple and many fast and robust methods exist for this purpose. Dense matching method involves matching each image pixel. The problem here is much more difficult. Research has been done towards combining the area-based and feature-based method resulting in hybrid methods.

The basic idea behind combining these methods is to integrate the information of the feature points in area based matching to produce a more efficient and accurate result. However, each of the individual methods has advantages and disadvantages.

3.1 Sparse-based Techniques

Although area-based approach algorithms attempt to provide dense depth data, this approach still suffers from the expensive computation cost. While most feature-based stereo matching systems are not restricted to using only a specific type of features, instead, a collection of feature types is incorporated. Since stereo vision involves extracting three dimensional data from the scene, the features which are useful in the stereo sense are features which describe the underlying 3D structures of the scene.

The first approach that raised the idea of using feature information is proposed by Barnard [4]. First, a set of candidate matching points (spots and corners) are selected independently in each image. After two sets of candidate points are found, possible matches are then constructed based on SSD correlations. This method does not aim at achieving a dense matching result but it integrates feature information other than only similarity measurement.

Lim and Binford [14], on the other hand, used a hierarchy of features varying from edges, curves, to surfaces and bodies (2-D regions) for high-level attribute matching. Pixel intensity, edge, and other surfaces are used to provide an over determination for matching.

Hoff and Ahuja [12], proposed a method that integrates feature matching, contour detection and surface interpolation. An estimation of the disparity map is used first to estimate the search area for matching. This estimate is done by comparing big windows of the correlation template and then, edges are extracted by an edge operator. The matching process is an integration of matching and interpolation based on the edge information. Matching and interpolation are done by fitting planar patches and finding the occlusion and rigid contours. Furthermore, the parameters will be changed successively by the edge operator. Fitting planar can be viewed as the matching of all points in a small local area and then matching and interpolation can be performed until the final result is reached. Although, this method achieves a dense disparity by combining interpolation and matching process, it is difficult to implement and computationally very expensive.

For instance, the system proposed by Weng [24] combines intensity, edges, and corners to form multiple attributes for matching. Edges and corners are blurred to different resolution levels to provide information needed for matching. Matching is done in a coarse to fine style in order to cope with large disparity and achieve final result. The epipolar constraint is not enforced as the approach is mainly aimed at matching non rigid scenes where the epipolar geometry does not hold. This method only gives a sparse matching result of a set of feature points, at lower cost.

3.2 Dense Matching Techniques

3.2.1 Exhaustive Search Approach

The computation required for matching a pixel to pixel in an unconstrained manner is not feasible and not reliable. Applying the epipolar constraint in matching can effectively reduce the search from a 2D image to a 1D epipolar line. Therefore, the epipolar constraint the match of an image pixel in the first image to lie on the same corresponding epipolar line in the second image. The match is the pair of the two image pixels with the best correlation score.

Later, this simple approach is modified by integrating interest points in each image, then attempt to establish the matching Zhang [26] on these points. The disparity estimate is based on the disparity of the matched interest points. Therefore instead of searching through the entire epipolar line for the candidate match, the search area uses the disparity information of the closest interest points.

Using epipolar constraints alone results in a high computational time. While the results of integrating interest points (Jin [13]), compared to the exhaustive search show significant improvements in CPU time and matching quality but the disparity guess based on image interest points is not always a good guess because given a pixel located on an object, the closest interest point to this pixel may be actually located on a different object and therefore, it might have a different depth in the scene.

3.2.2 Dynamic Programming

The main approach of dynamic programming is to organize an optimization problem to reduce redundant calculations using the solutions to sub-problems. Thus, in the case of matching, Ohta and Kanade's dynamic programming approach [20] assumes epipolar lines coinciding with the horizontal scan lines and they depend mainly on the ordering constraint in the matching process. Lloyd [15] presents a two stage method that produces a set of candidate matches, then using the continuity constraints to choose the best among the candidates.

Although dynamic programming approach decreases computation in the matching process, the reduction of calculation comes at a cost. If calculation is reduced, a large amount of data needs to be stored. A problem will occur when using large images.

3.2.3 Simulated Annealing

Barn [3] attempted matching the parallel stereo images using simulated annealing by assuming epipolar lines coinciding with the horizontal scan lines and the two retinal planes are horizontally displaced and are coplanar in space, and the two cameras have identical focal length. He defined an energy function E_{ij} as:

$$E_{ij} = |I_L(i,j) - I_R(i, j + D(i,j))| + \lambda|\delta D(i,j)| \tag{3.1}$$

where I_L denotes the intensity value of the left image at the $i-^{th}$ row and $j-^{th}$ column and I_R denotes the intensity value of the right image at the same row but at the $k-^{th}$

column; $D(i, j)$ is the disparity value (or horizontal shift in this case) at the ij-position of the left image. The above is clearly a constrained optimization problem in which the only constraint being used is a minimum change of disparity values. This constraint is commonly known as the continuity constraint.

Robe [22], later incorporated the use of a multiresolution scheme together with a smoothness constraint similar to that of Barn [3] into the constrained optimization process. In addition to the horizontal shift of corresponding pixels, they also allowed the corresponding pixels to undergo vertical shift (i.e. disparity in the vertical direction), so their matching method is not restricted to only parallel stereo images. The energy function to be minimized, as expected, is more complicated than the one given above.

The advantage of this approach is that a dense disparity map, and consequently a dense depth (or range) map is output. Unfortunately, like all constrained optimization problems, whether the system would converge to the global minima is still an open problem, although, as reported by Robe [22], the multiresolution scheme, to a certain extent, helped speed up convergence and avoid local minimal.

3.2.4 Window-Based Method

Okutomi [21] propose a stereo matching method using an adaptive window. This approach is to match only those regions in the images that contain high variation of intensity values in the horizontal, vertical, and diagonal directions. However, the model of the disparity map must have been previously determined to estimate the optimal window size at each position. The results obtained by the method are better than any correlation-based stereo matching method using a fixed size window, demonstrating the advantage of the adaptive window.

The problem associated with this window-based approach is that the size of the correlation window must be carefully chosen. If the correlation windows are too small, the intensity variation in the windows will not be distinctive enough and the result will contain many false matches. If they are too large, resolution is lost, and no good match may found, since neighboring image regions with different disparities will be combined in the measurement.

To determine the optimal window size without any disparity model, stereo images and their generated disparity maps are evaluated, the window size that optimizes the evaluation is selected. A generated disparity map is evaluated using the compatibility between corresponding points and also map continuity (Marr [19]). Also some genetic algorithm (GAs) is applied. For instance, Saito [23] employs several disparity maps that are generated by SSD correlation using different window sizes, then the optimal disparity map is determined by combining these maps using a genetic algorithm.

The problem of adaptive window approach is that, if various window sizes are used to find correspondence, then for each point, several candidate values of disparity exist. Therefore, the number of candidate disparity maps is huge.

3.2.5 Iterative Matching Technique

The goal is to find a pixel mapping that represents a globally optimal solution. Based on the pixel's neighbors, and a set of constraints, the pixel is matched in an iterative process. Relaxation labeling is one such approach. For each match pair (p, p') in the image, the probability of the pair being the proper match is calculated. The probabilities are computed from similarities in the value surrounding the match points. Then these probabilities are iteratively updated, based on the probabilities of neighboring pixels and a set of constraints, until a steady state is achieved. For example, Christmas *et. al.*, [6] define a probabilistic method that integrates contextual information using a Bayesian framework. This contextual information is used as evidence to provide a formula that prescribes in a unified and consistent manner how unary relation measurements relating to single entities, binary relation measurements relating to pairs of objects should be brought to bear on the object labeling problem.

A recent energy-based method, L. Alvarez [1] solves the correspondence problem as a minimization problem that preserve discontinuities at object boundaries. This iterative approach is far from straightforward to implement.

3.2.6 Region Segmentation and Matching in Stereo Images

Boufama and O'connell [5] used a method to simultaneously achieve segmentation and dense matching in a pair of stereo images. The method starts with a rough estimate of a plane, defined by three points. In particular, segmentation is a geometry-base instead of color based. This method based on geometry, uses correlations only on a limited number of key points in contrast to previous methods that are based on similarities or correlation techniques. Segmentation is the process of identifying regions of similarity in the image. When segmentation is achieved, matching can rely on its result and can be easier to solve. The approach of segmentation is intended for scenes with planes, such as indoor scenes and outdoor city scenes.

Although the method used on real images proved its capability and performance, more work needs to be done at the plane identification level.

Iterative techniques have good success in finding the globally optimal solution but their main drawback is that the time required to reach this optimal solution is quite large.

3.3 Conclusion

Search methods, such as dynamic programming, simulated annealing and window-based, aim to reduce the size of the candidate set. The calculation of correlation methods is computationally intense and not feasible for real-time dense matching while most existing hybrid dense matching algorithms are not practical and time consuming. Improvements must be made when possible. For our work, we will consider a pair

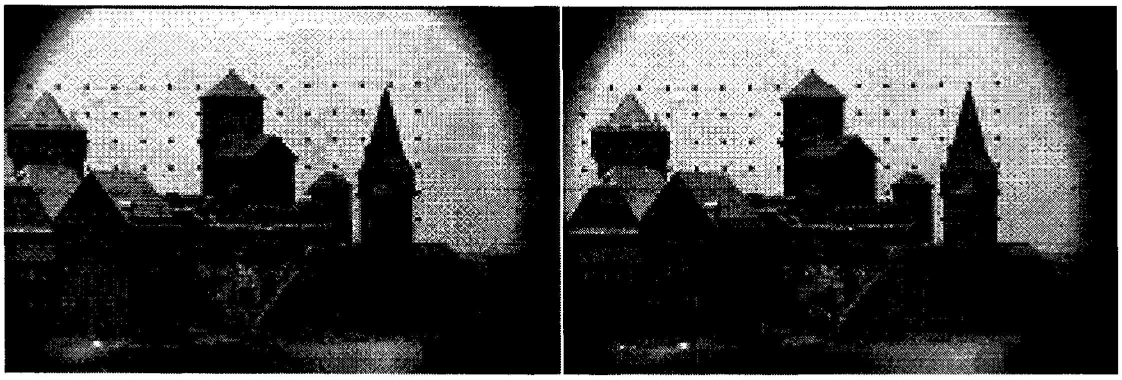


Figure 3.1: Stereo image pair: Castle scene(size 576x384).

of 2D uncalibrated images (Figure 3.1), of a 3D object are taken from two distinct viewpoints and undergo a horizontal motion. Our intend is to design a hybrid matching algorithm that can reduce the process time of a dense matching and enables simple and straightforward implementation.

Chapter 4

Contribution To Fast Dense Matching

Our contribution to dense matching employs a hybrid method of both the feature-based and area-based strategies. Chapter 3 have shown the drawbacks of the previous work, where the number of mismatches and the computation time is high. Since dense matching methods are based on correlation techniques to evaluate regions similarity, these methods are too slow because correlations are calculated intensively in the matching process. Moreover, although some methods have attempted to integrate image feature information in the dense matching of uncalibrated images, most of these methods are not practical.

Therefore, we need to consider more reliable image features in our approach. Our approach is to consider important type of image features: the edges. Edges have advantages over interest points since edges reveal a lot of information on the image and allow us to locate a possible abrupt disparity change. Locating edges within an image is an important part of any low-level computer vision system.

Match n points in a non-edge segment $AB[n]$;
begin
for $(k = 0, k < N; k^{++})$
if match(p != 1) , 1 for edge pixel and $p(x, y)$ ϵ edges in leftImage;
continue;
else
match(p)= $maxZNCC(p, q)$, $q(x\prime, y\prime)$ ϵ search area in rightImage;
count++;
if ($count$ == 2), match two segments $AB[n]$ and $A\prime B\prime[n\prime]$;
$t = (x_B - x_A)$;
$\lambda = \frac{(x\prime - x_0\prime)}{t}$;
for $(j = x\prime_o, j < x\prime, j^{++})$
match $AB(p(x_A)) = A\prime B\prime(p\prime(x_{A\prime} + t \times \lambda))$, lambda the scale factor of two segments;
x^{++};
End

Table 4.1: Pseudo code for our hybrid method using an algebraic method.

4.1 A new hybrid algorithm

Our method is a hybrid of both the sparse-based and dense matching strategies. A flow diagram of our approach is shown in Figure 4.1 is to present a fast dense matching algorithm which integrates the edge regions and non edge regions of images. First, edges are extracted from the left image which is divided into two sets of regions, edge and non-edge regions. Then, we match all pixels belonging to the edge regions of the left image to their corresponding pixels in the right image using correlation and epipolar constraint. Finally, non edge regions are matched using another algorithm that uses an algebraic method to restrict the search area. A final dense matching result is achieved by combining the matching of edge and non-edge areas. Table 4.1 gives the pseudo-code description for the algorithm.

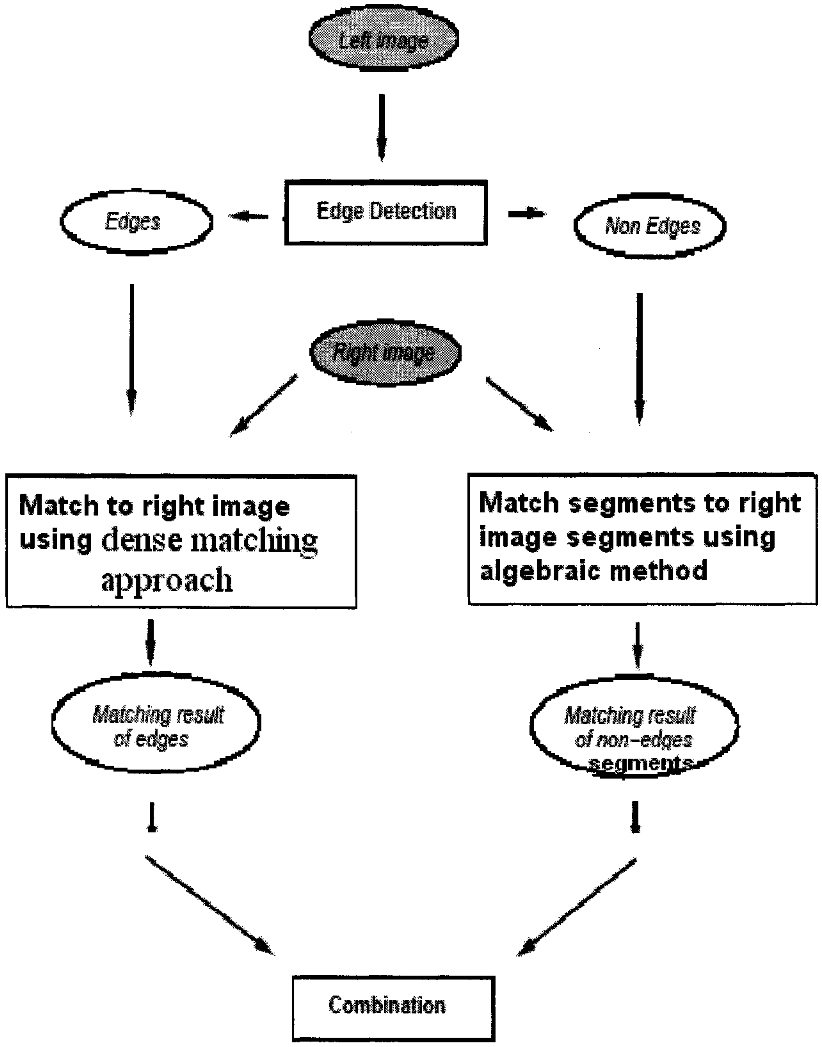


Figure 4.1: Flow diagram.

4.2 Matching Edge Area

An edge occurs where there is a discontinuity in the intensity function. Our work involves two steps: extracting edges from the left image and then using a threshold to obtain more real edges. The edge detector we use in our work is a Sobel edge detector.

- Sobel Edge Detection: it returns edges at those points where the gradient of the intensity image is maximum. Two 3x3 convolution masks are applied to each pixel, one color at a time; one with a horizontal trend and one with a vertical trend. Result of each convolution is treated as a vector representing the edge through the current pixel. If magnitude of the sum of these two orthogonal vectors is $>$ than a specified threshold, the pixel marked in black as an edge, otherwise, the pixel is set to white. Also, Sobel operator smooth the image in horizontal and vertical directions reducing the effect of random noise for corrupting the estimate of edge contrast [8, 7].

- Thresholding: thresholding the output of the detected edges to select the strongest edges and set everything else to white. After setting a threshold, in a single pass, each pixel in the image is compared with this threshold. If the pixel's intensity is equal or higher than the threshold, the pixel is set black in the output. If it is less than the threshold, it is set to white. The importance of adapting a threshold is that it shows comparable results using the matching process.

Figure 4.2 and Figure 4.3 show the output of Sobel edge detector extracted from the left image of the castle scene and the edge area we extracted after Sobel edges using a threshold =20 after comparing differing threshold value in the left image to detect more edges. As stereo matching requires a pixel to pixel comparison, it is necessary to employ a correlation measure to evaluate the fitness of potential matches in order to match the edge regions. Therefore, all the intersections between the epipolar line and edges in the first image are detected and finally all the edges on all the epipolar lines are matched using area matching approach. We enforce the epipolar constraints as the

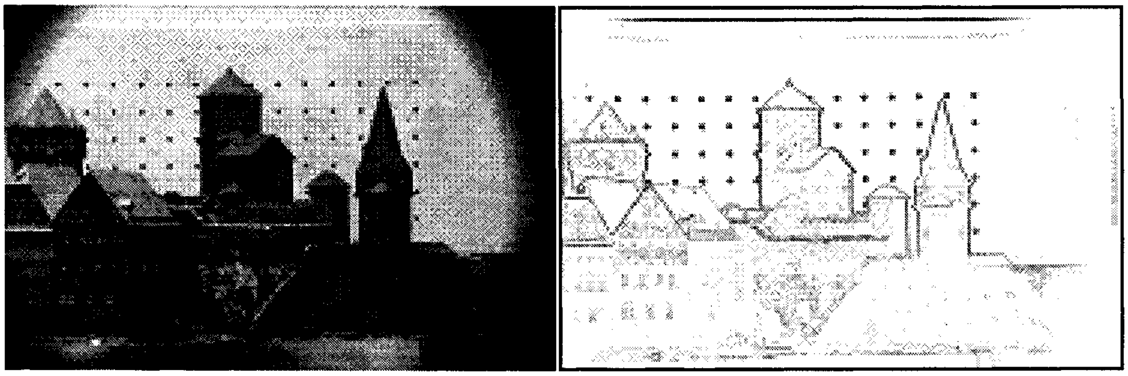


Figure 4.2: Sobel edge detection for the left image.

images undergo horizontal motion. Considering a point $p = (x, y)$ in the first image and its corresponding point $p'(x', y')$, $y = y'$ as there is no vertical parallax in parallel images.

We choose ZNCC in our work as it is a more robust correlation function than traditional ones, (see section 3.). The ZNCC of each pixel is calculated and the pixel associated with the highest score is taken as the match. As ZNCC gives a measure on an absolute scale range of [-1,1] of the degree of similarity between two areas-based on the pixels gray values, the higher the similarity of these neighborhoods, the better the correlation score.

As we mentioned before, our image pair undergoes a horizontal motion (Figure 4.4), the epipolar lines coincide with the horizontal scanlines because the cameras are parallel. Furthermore, the corresponding points in the right image must then lie on the same horizontal scanline and the epipolar reduces the correspondence problem to a line search.

Therefore, for every pixel of interest in the left image, we search along the epipolar line for the corresponding pixel. Mismatch will be unlikely as the epipolar line is applied as well multiple matches are very unlikely as the images have a horizontal motion and the corresponding matches in the right image will lie on the same corresponding

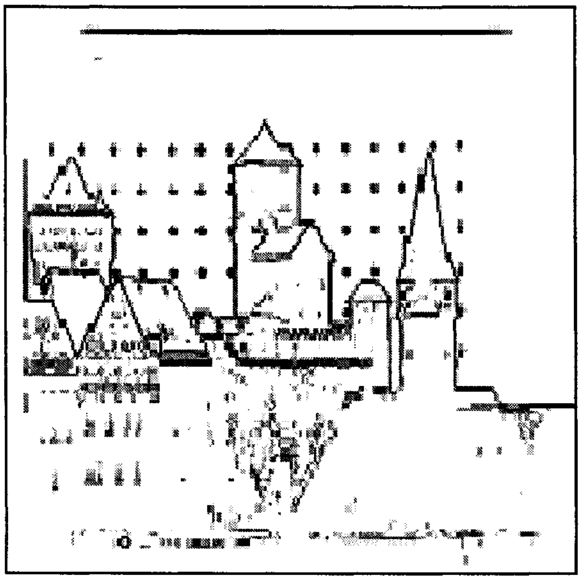


Figure 4.3: Thresholding.

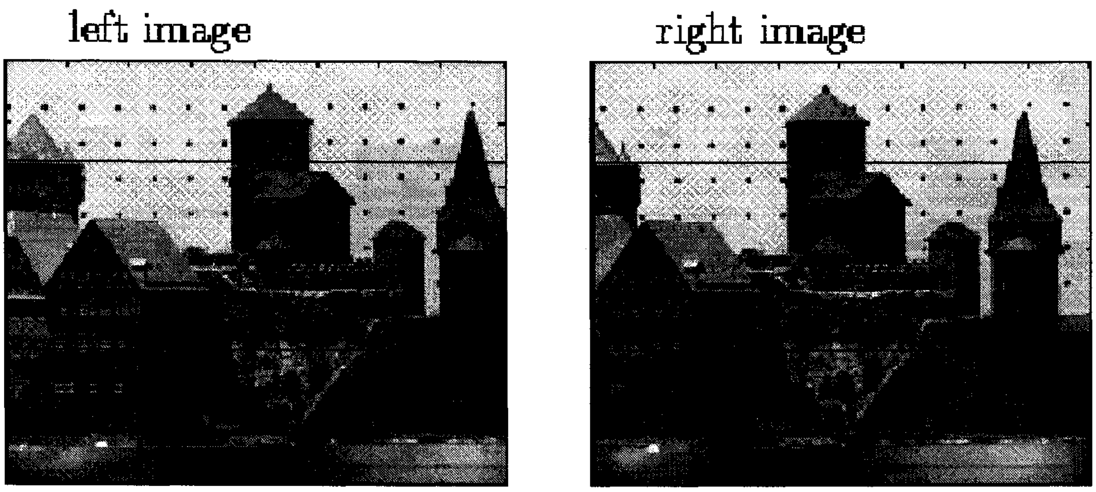


Figure 4.4: The matching points lie on the same horizontal scanline.

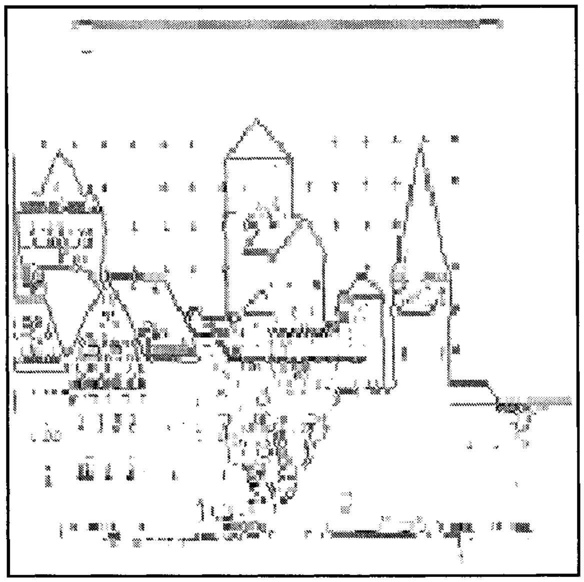


Figure 4.5: The matching result of edge areas.

horizontal line of the left image. Since edges' pixels represent a fraction of the whole image pixels, using correlation-based method is not costly in terms of CPU-time. Figure 4.5 shows the result of matching edge areas that we have obtained.

4.3 Matching non-Edge Area

4.3.1 Identifying non-Edge Segment

When edge areas are extracted, the non edge areas are obtained by subtracting edge regions from the original image. Since the matching is carried along epipolar lines and disparity continuity is preserved at the object, we define a non-edge segment as a sequence of non-edge pixels on an epipolar line delimited by two edge areas. Figure

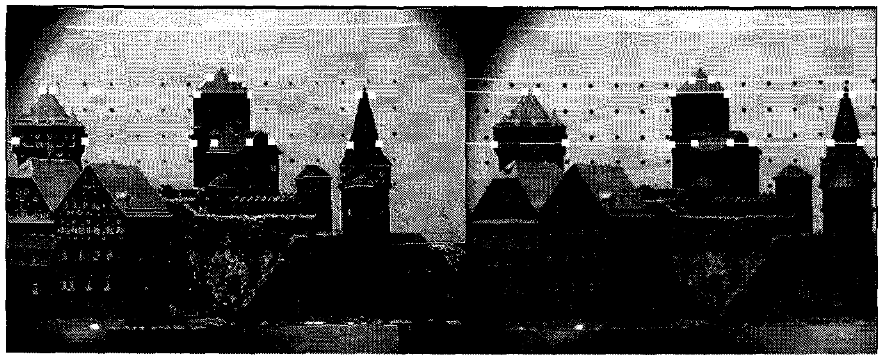


Figure 4.6: Our example of non-edge segments.

4.6 shows an example of non-edge segments on three corresponding epiolar lines of the castle scene. Consider non-edge segments along epipolar lines that cross the whole image. One can note that we have different non-edge segments delimited by two edge areas marked in white color on each epipolar line.

4.3.2 Matching non-Edge Segments using an Algebraic Method

Matching of image's non-edge areas can be transformed into the matching of all the non-edge segments on all the epipolar lines since every pixel from the non-edge areas falls into a certain non-edge segment. Since the non-edge areas consist mainly of smooth surfaces without abrupt disparity changes, all the following constraints are obtained implicitly: order, continuity and uniqueness. As continuity constraint states, discontinuity changes only occur at edges and by definition a non-edge segment has no edge pixels. Therefore the search for the candidate match is limited to a restricted area.

Consider matching process of a non-edge segment $AB[N]$ lying on the epipolar line in left image and $A'B'[M]$ the corresponding segment lying on the corresponding epipolar line in the right image. $AB[N]$ and $A'B'[M]$ denote the N and M pixels of these segments respectively.

If we have a matched pair of pixels such as:

$$AB[i] \leftrightarrow A'B'[j]$$

where we refer to $AB[i]$ as the *reference point*, its match $A'B'[j]$ the reference match, and $(AB[i], A'B'[j])$ the reference pair,
then

$$AB[i+1] \leftrightarrow A'B'[j+1].$$

As the order of matching is preserved along the epipolar line, the pixel p on the right of the reference point must be matched to a pixel on the same side of the reference pair, while ignoring the scaling that occurs in the image.

On the other hand, considering the scale change of the image, we need to add the scale factor λ and the offset t. Here, λ is the ratio of the distance between two reference edge matches of a segment located on the corresponding epipolar line in the right image divided by the distance of the two reference edges points of a segment located on the epipolar line in the left image, and where t is the distance between two reference points in the left image.

As shown in Figure 4.7, let (x_A, x_B) be the X-axis coordinates of the non-edge segment AB of pixel i and $i\prime$ successively in the left image and let $(x_{A\prime}, x_{B\prime})$ be the X-axis coordinates of the corresponding non-edge segment $A'B'$ of pixel j and $j\prime$ successively in the right image knowing that we have parallel images. To match a certain pixel within the same edge segment, we take into consideration the scale factor λ and t.

Therefore, the scale factor is defined as follows:

$$\lambda = \frac{(x_{B\prime} - x_{A\prime})}{(x_B - x_A)} = \frac{(x_{B\prime} - x_{A\prime})}{t} \Rrightarrow t_{AB} \times \lambda = (x_{B\prime} - x_{A\prime}) \tag{4.1}$$

Within a non-edge segment, we conclude from equation (4.1), that the search area for a candidate match is restricted to the offset t_{AB} of two reference points in the left image times the scale factor λ plus the left neighborhood of the candidate match such as

$$x_{B\prime} = x_{A\prime} + t_{AB} \times \lambda \tag{4.2}$$

which is a linear equation applies to all segments in the images and where

$$t_{AB} = (x_B - x_A) \tag{4.3}$$

Now, let a reference point p with ordinate x_p vary then its reference match $p\prime$ with ordinate $x_{p\prime}$ varies always restricted to the linear equation below:

$$x_{p\prime} = x_{A\prime} + t_{AP} \times \lambda \tag{4.4}$$

and where

$$t_{AP} = (x_p - x_A) \tag{4.5}$$

It must be noted if $\lambda > 1$, the non-edge segment $A'B'$ has been stretched, if $\lambda < 1$, the non-edge segment $A'B'$ has been shrunk with respect to left image. However, For each non-edge-segment, the scale factor λ and the offset t are constant, only $x\prime$ varies as x varies. And since λ represents a small neighborhood around the reference match, the possible match of pixel p located around the right neighborhood of pixel i on the left epipolar line must be around the right neighborhood of j in the right image, so the matching relation becomes:

$$AB[i+1] \leftrightarrow A'B'[j+1+t \times \lambda].$$

As disparity varies smoothly on object surfaces and sharp changes of disparity occur at object boundaries, the offset t and the scale factor within two non-edge segment reference pair provides an accurate base for matching other pixels in the same segment. Thus, we can conclude that the algebraic method using the scale factor of two matching non-edge segments pair verifies the epipolar, continuity and order constraints. Now this

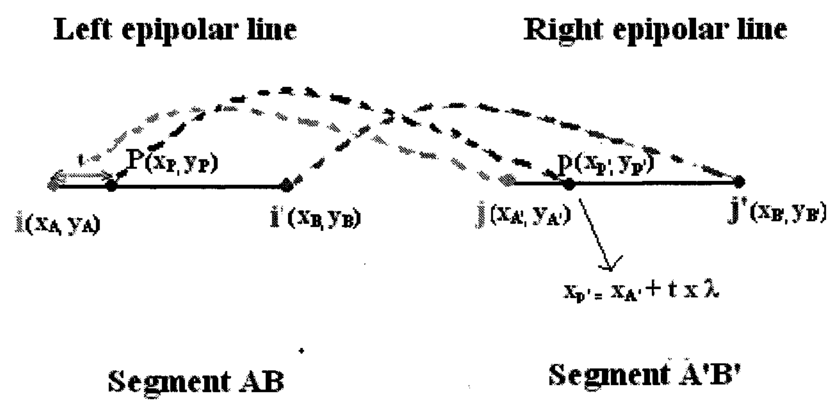


Figure 4.7: Using algebraic method to match non-edge segments.

method can be applied to the n^{th} neighbor of the reference point within the same non-edge segment. So the matching relation becomes for a matched pair:

$$AB[i] \leftrightarrow A'B'[j]$$

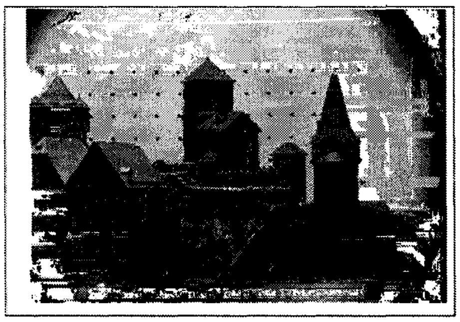
as follows:

$$AB[i+n] \leftrightarrow A'B'[j+n+t \times \lambda],$$

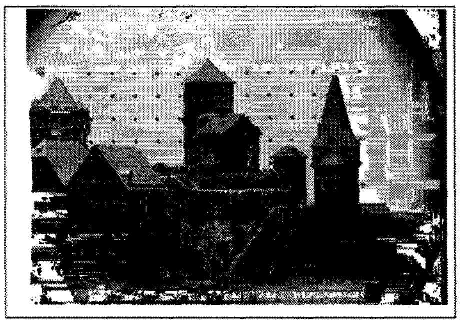
where $AB[i+n]$,$AB[i]$ $\in$ same non-edge segment.

4.3.3 Experimental Results

We have tested our method on three different scenes: Castle scene (Figure 3.1), Head scene (Figure 4.10) and Meter scene (Figure 4.13). In our experiments, the matching is carried out from the left to right image, we display the reconstructed right image from the matching result as the output.



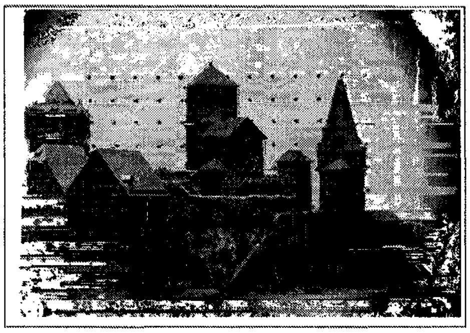
Correlation window size 15x15



Correlation window size 13x13

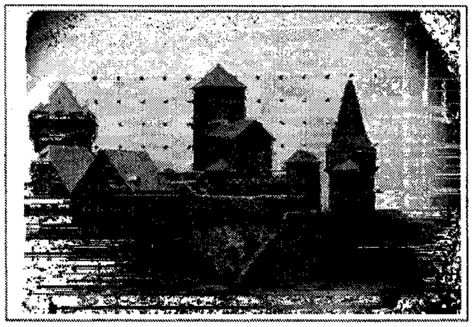
Correlation window size 11x11

Figure 4.8: Matching results using hybrid approach for Castle scene 1.



Correlation window size 9x9

Correlation window size 7x7



Correlation window size 5x5

Figure 4.9: Matching results using hybrid approach for Castle scene 2.

Figure 4.10: Stereo image pair: Head scene (size 456x343).

- Results on the Castle scene: the results of using the hybrid method is shown in Figure 4.8 and Figure 4.9. The figures show reconstructed images using different correlation window sizes for the Castle scene. The results show an improvement non-edge segments were detected and matched. The quality of the reconstructed image is slightly improved using a bigger correlation window. A few gaps of white spots appeared due to pixels that were not matched because their low correlation scores, or a non-edge segments were not detected and therefore don't have a match. But the original objects is greatly recognized and the hybrid method demonstrates its capability of preserving discontinuity around edges. Furthermore, our hybrid matching produced good matching results with low time cost.

- Results on the Head scene and Meter scene: more experiments have been performed on the head scene and Meter scene. The results are shown in Figure 4.11 and Figure 4.12, Figure 4.14 and Figure 4.15. The Head scene is an indoor scene with objects of more depth differences. The results is considerably acceptable to both matching accuracy and efficiency. Good results obtained with low time consuming. As well, the likelihood of a mismatch to happen is low because

Correlation window size 15x15

Correlation window size 13x13

Correlation window size 11x11

Figure 4.11: Matching results using hybrid approach for Head scene 1.

Correlation window size 9x9

Correlation window size 7x7

Correlation window size 5x5

Figure 4.12: Matching results using hybrid approach for Head scene 2.

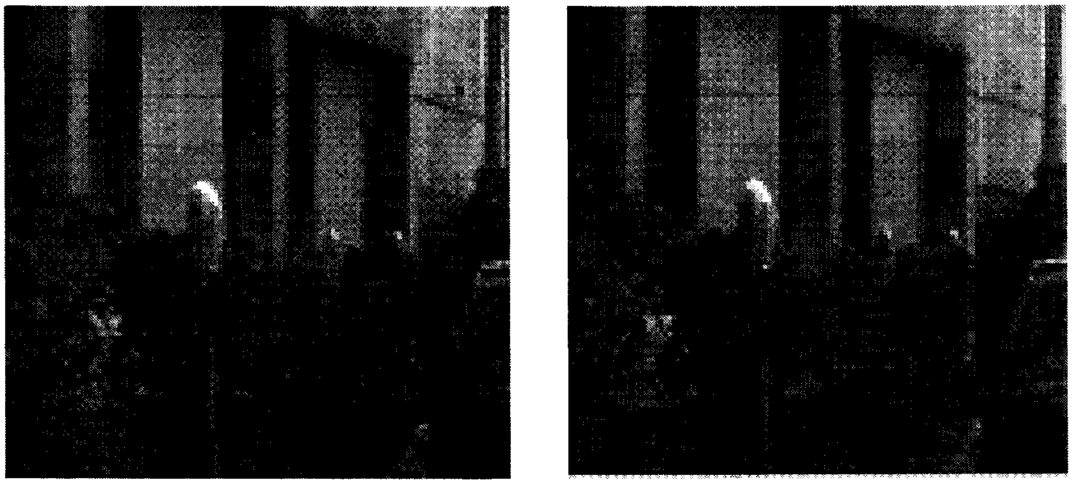


Figure 4.13: Stereo image pair: Meter scene (size 308 X 288).

Castle scene(image size 476 X 384)

Correlation window size	15X15	13X13	11X11	9X9	7X7	5X5
CPU Time	85s	68s	55s	37s	33s	17s

Head scene (image size 456 X 343)

Correlation window size	15X15	13X13	11X11	9X9	7X7	5X5
CPU Time	50s	45s	40s	30s	24s	15s

Meter scene (image size 308 X 288)

Correlation window size	15X15	13X13	11X11	9X9	7X7	5X5
CPU Time	27s	17s	16s	14s	12s	9s

Table 4.2: Processing time using our hybrid approach for the three stereo images.

Correlation window size 15x15

Correlation window size 13x13

Correlation window size 11x11

Figure 4.14: Matching results using hybrid approach for Meter scene 1.

Correlation window size 9x9

Correlation window size 7x7

Correlation window size 5x5

Figure 4.15: Matching results using hybrid approach for Meter scene 2.

the scale factor within two non-edge segments points provide an accurate base for matching other pixels in the same segment. This indicates that using the algebraic method is a good choice for stereo matching.

Moreover, Table 4.2 shows the computation time was very low noting that the programs are run on a Pentium 4 machine. One can note that the correlation window size has a slight effect on the CPU-time when matching non-edge segments. However, our hybrid method combining feature-based method and area based method shows its advantages over traditional methods by improving the matching quality and considerably increasing the matching speed, noting that our results don't include interpolation which is a procedure that estimates missing values within an area of known values.

4.4 Conclusion

In this chapter, we proposed a fast algorithm for dense matching of uncalibrated images. As edges reveal a lot of information on the image and indicate the locations of possible abrupt disparity changes, image is segmented into two parts: edge regions and non-edge regions. The edge regions were matched using correlations and the epipolar constraint only. This has proven to be sufficient to obtain very reliable matching results on the edge regions and to speed-up the processing time. On the other hand, for the rest of the image which is composed on non-edge regions, the matching approach using an algebraic method to match non-edge segment was used. Thus an accurate searching strategy is achieved. The experimental results have shown matching quality and time consuming using our method comparing to the correlation functions, as ZNCC is one of the most CPU-time expensive. Our hybrid method has been tested on several indoor and outdoor scenes and our experimental results demonstrate its capabilities.

We conclude that the image features in process are reliable in the matching process. Therefore, we believe that our method is time consuming and easy to implement.

Chapter 5

Conclusion

Our contribution of this thesis was to implement a fast matching method for dense matching of uncalibrated images. The drawbacks of the traditional method is its high computational time and poor matching. Integrating image features such as edges in the matching process are more reliable and efficient.

Experimental results introduced the advantage of our hybrid method that combine sparse matching technique and dense matching techniques. Such advantages included a fast dense matching and robustness against image noise, as well mismatches was reduced comparing to traditional methods. We have used an algebraic method to match non-edge segments and that has been proven to Work for area with no texture.

As our hybrid method demonstrates its capability of preserving discontinuity around edges, the quality of the reconstructed images can be even better by detecting more edges before the matching process taking place.

5.1 Future Work

Matching features across images is ill-defined problem and although much work has been done in this area, most of it is not robust by any means.

- Although the quality of matching edge-areas is good, it can be made even better by adding constraints on the search areas for a better dense matching. Like possibility of limiting the search to the edge-area in the second image.

- Our work used stereo images with horizontal motions. But in general, epipolar lines are not aligned with coordinate axis and are not parallel. Such searches are time consuming since we must compare pixels on skew lines in image space. These types of algorithms can be simplified and made more efficient if epipolar lines are axis aligned and parallel. This can be realized by applying image rectification.

Bibliography

[1] L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert. Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach. Technical Report 3874, Inria, (2000).

[2] P. Aschwanden and W. Guggenbuhl. Experimental results from a comparative study on correlation-type registration algorithms. Robust Computer science, pages 268-282. Wichmann, (1992).

[3] S. T. Barnard, "Stereo matching by hierarchical, microcanonical annealing". *In Proceedings of International Joint Conference on Artificial Intelligence,* pages 832–835.(1987).

[4] S. T. Barnard and W.B. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*2(4):333-340, 1980.

[5] B. Boufama and D. O'Connell: Region Segmentation and Matching in Stereo Images. *In the 16th International Conference on Pattern Recognition,* pages 631–634.(2002)

[6] W. J. Christmas, J. Kittler, and M. Petrou, Probabilistic relaxation for matching problem in computer vision. *In Proceedings of Forth International Conference on Computer Vision,* pages 666–673, Gemany.(1993).

[7] E. Davies Machine Vision: Theory, Algorithms and Practicalities, *Academic Press*,Chap 4, (1990).

[8] R. Gonzalez and R. Woods Digital Image Processing, *Addison-Wesley Publishing Company*, Chap7, (1992).

[9] P.Gross. How to use the cross ratio to compute invariants from two images. *In J. Mundy, A. Zisseman, and D. Forsyth, editors, DARPAESPRIT workshop on Applications of invariants in computer vision,* pages 107-126 in Lectures Notes in Computer Science, Azzores, Portugak, Springer-Verlag, 1993.

[10] Hanson, N. R.*Patterns of Discovery.* Cambridge: Cambridge University Press, (1958)

[11] R. Hartley. In defence of the eight-point algorithm. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580-593, (1997).

[12] W. Hoff and N. Ahuja. Surfaces from stereo:mIntegrating feature matching, disparity estimation, and contour detection.*IEEE Transactions on Pattern Analysis and Machine Intelligence,* 16(9): 920-932, (1989).

[13] H. JIN. Dense Matching of uncalibrated Images for stereo vision. Master of science, University of Windsor, (2001).

[14] H. S. Lim, and T. O. Binford, Stereo Correspondence: A Hierarchical Approach, in *'Proc. Image Understanding Workshop',* Vol. 1, pp. 234-241, (1987).

[15] S.A. Llyod, Stereo matching using intra- and inter-row dynamic programming. *Pattern Recognition Letters,* 4:273-277, (1985).

[16] H.C.Longuet-Higgens, A computer algorithm for reconstructing a scene from two projections. *Nature,* 293(10): 133-135, September (1981).

[17] C. Loop, Z. Zhang, Computing rectifying homographies for stereo vision. *In Proceedings of IEEE on Computer vision and Pattern recognition* ,Vol.1, pp. 125 - 131, Colorado, USA, (1999).

[18] Q.Luong and O.Faugeras, Calibration, scene motion and structure recovery from point correspondences and fundamental matrices. *International Journal of Computer vision,* (1996).

[19] D. Marr, and T. Poggio, A Computational Theory of Human Stereo Vision, in *'Proc. Royal Society of London',* Vol. 204 of B, pp. 301-328, (1979).

[20] Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI-7(2):139-154, (1985).

[21] M. Okutomi and T. Kanade, A stereo matching algotithm with an adaptive window theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9): 920-933, (1994).

[22] L. Robert, R. Deriche, and O. D. Faugeras, Dense Depth Map Reconstruction Using Multiscale Regularization, *in 'Proc. International Conference on Image Processing',* Singapore, pp. 123-127.(1992).

[23] H. Saito, and M. Mori. Application of genetic algorithms to stereo matching of images. *Pattern Recognition Letters,* 16(1): 815-821, (1995).

[24] J. Weng, N. Ahuja, and T. S. Huang, Matching two perspective views. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 14:806–825.(1992).

[25] J. Weng , Y. Liu, T. S. Huang, and Ahuja, N. Estimating Motion/Structure from Line Correspondences: A Robust Linear Algorithm and Uniqueness Theorems, in *'IEEE Proc. Computer Vision and Pattern Recognition',* Ann Arbor, Michigan, pp. 387-392,(1988).

[26] Z.Zhang. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence* 78:87-119, (1995).

VITA AUCTORIS

NAME	Rabihaa Zein
PLACE OF BIRTH	Niha, Lebanon
EDUCATION	Beirut University College, Beirut, Lebanon 1985-1990 B.SC.
	University of Windsor, Windsor, Ontario 2002-2005 M.Sc.